

## Language Shift by Indigenous Population: A Model Genetic Study in South Asia

Gyaneshwer Chaubey<sup>1,2\*</sup>, Mait Metspalu<sup>1</sup>, Monika Karmin<sup>1</sup>, Kumarasamy Thangaraj<sup>2</sup>, Siiri Rootsi<sup>1</sup>,  
Juri Parik<sup>1</sup>, Anu Solnik<sup>1</sup>, Deepa Selvi Rani<sup>2</sup>, Vijay Kumar Singh<sup>2</sup>, B. Prathap Naidu,  
Alla G. Reddy<sup>2</sup>, Ene Metspalu<sup>1</sup>, Lalji Singh<sup>2</sup>, Toomas Kivisild<sup>1,3</sup> and Richard Villems<sup>1</sup>

1. Department of Evolutionary Biology, Institute of Molecular and Cell Biology, University of  
Tartu and Estonian Biocentre, Tartu, Estonia

2. Centre for Cellular and Molecular Biology, Hyderabad, India

3. Leverhulme Centre of Human Evolutionary Studies, The Henry Wellcome Building,  
University of Cambridge, Fitzwilliam Street, Cambridge, CB2 1QH, UK

**KEYWORDS** Mushar; language shift; mtDNA; Y-chromosome

**ABSTRACT** Language shift is a phenomenon where a new language is adopted by a population with virtually no influence on its genetic makeup. We report here the results of a case study, carried out on the Mushar populations, which is thought to have undergone language shift from Munda (an Austro-Asiatic language) to Hindi (an Indo-European language). We compared the mtDNA and Y-chromosomal phylogenies of this population with those of the neighbouring Indo-European and Austro-Asiatic speaking populations, standing at similar social status. The results revealed much closer genetic affinity of the Mushar people to the neighbouring Austro-Asiatic (Mundari) populations, than to the neighbouring Hindi-speaking populations. This example shows that the language shift as such is not necessarily a signal for a rapid genetic admixture, either maternally or paternally.

### INTRODUCTION

The out of Africa migration has led the first footprints of anatomically modern humans (AMHs) to South Asia many tens of thousands of years ago. In addition to this, the *in situ* origin of numerous deep rooting indigenous haplogroups (Kivisild et al. 2003; Palanichamy et al. 2004; Sun et al. 2006; Thangaraj et al. 2006) and the development of a complex caste system has made it highly diverse, both genetically and socially. There are several controversial issues *viz.* the beginning of caste system and origin and migration of different language groups that remain yet to be resolved (Kivisild et al. 2003; Metspalu et al. 2004; Sahoo et al. 2006; Sengupta et al. 2006; Chaubey et al. 2007; Kumar et al. 2007). The Indian subcontinent is linguistically divided into four major families: Indo-European, Dravidian, Austro-Asiatic and Tibeto-Burman. Indo-European is the most widely spread family, which covers North, Central and Western parts

of India. Dravidians are mainly confined to South India with some exceptions like Kurukh and Gondi speaking people, living in central and eastern parts of India. Austro-Asiatic speakers are dispersed mostly in the central and eastern parts of the country while the Tibeto-Burman speakers are concentrated in the northeastern part of the subcontinent. The origin and dispersal of these language families in South Asia remains still as enigmatic.

The caste system of South Asia is maintained by several traditional rituals and social practices, including strict endogamy. Other than the scheduled castes (SC) population, the scheduled tribal population (ST) represent the historically most backward and disadvantaged groups in the highly ripped Indian caste society. While the social, economic and educational deprivation of these groups has been in the past a common and unifying characteristic, each group had and still largely has its own particular rituals that distinguishes one from the other. In the past few years, genetic studies of mtDNA and Y chromosomal variation have provided a substantial understanding of the human origins and dispersal patterns in South Asia. (Kivisild et al. 1999, 2003; Cordaux et al. 2003; Palanichamy et al. 2004; Metspalu et al. 2004; Sun et al. 2006; Thangaraj et al. 2005, 2006; Sahoo et al. 2006; Sengupta et

\*Corresponding Author: Gyaneshwer Chaubey,  
Department of Evolutionary Biology, Tartu University  
and Estonian Biocentre, Riia 23 Tartu, 51010  
Estonia  
Telephone: (372) 7 375 052  
Fax: (372) 7 420 194  
E-mail: gyanc@ebc.ee

al. 2006 ; Chaubey et al. 2007; Kumar et al. 2007). Mitochondrial DNA analysis, in particular among the worldwide populations, has revealed continent-specific distributions of certain mtDNA lineage groups. (Kivisild et al. 2003, 2006; Metspalu et al. 2004; Torroni et al. 2006). Various studies on South Asian populations do not show any caste- or language-specific distribution of major ancient lineages (Kivisild et al. 2003; Metspalu et al. 2004; Sun et al. 2006). The internal structure of haplogroup M lineages, as revealed now by the analysis of complete mtDNA genomes in South Asia reflects their basically autochthonous origin (Metspalu et al. 2004; Sun et al. 2006; Thangaraj et al. 2006), thus confirming our earlier predictions (Kivisild et al. 1999, 2000, 2003).

Language shift is a process where a more prestigious, usually near-by language is adopted by a contemporary population, with only a minor contribution of genes from a source population (Diamond and Bellwood 2003 and references therein). It is a process whereby a community shifts to speaking another language. In such cases generally a group, for the sake of the social and economic status, tends to assimilate the language of the majority host population. Language shift may be also forced by ruling authorities. The loss of original language is usually gradual and extends over several generations. An example has been given by Diamond and Bellwood (2003) where the growing adoption of English as the language of government in Papua New Guinea (PNG), whose original populations nonetheless remain indigenous with negligible input of European genes. Historically, language shifts are common - e.g. in British Islands, in Roman-time Sardinia, in Armenia after the fall of Urartu, in Anatolia with the arrival of Turks (Cavalli-Sforza et al. 1994) all these very likely with an only a limited genetic change. Few other prehistoric examples are the adoption of Austronesian languages by some former speakers of Papuan languages in the western islands of Melanesia and by Agta Negritos in the Philippines (Kulick 1992, Cavalli-Sforza et al. 1994, Ross 1997). A prehistoric language shift from Uralic to Indo-European among the Baltic-speakers has been also hypothesized (Wiik et al. 2002).

There is a lack of detailed studies of language shifts in South Asia. Therefore, we investigate the correlation between language and genes in a population called Mushar which has, according to the earlier reports (Sharma 1998; Mishra 2001;

Sinha 2002), highly likely changed its language from Austro-Asiatic (Munda) to Indo-European in the recent past. To address this question we analyzed maternal (mtDNA) and paternal (Y-chromosome) lineages of the target population and compared the results with those of the neighbouring Austro-Asiatic and Indo-European populations.

## MATERIALS AND METHODS

### Subjects

The word Mushar is related to Bhojpuri word 'Mush' meaning mouse or rat and 'ahar' means eaters (Chaubey et al. 2007). They smoke into the rat holes in the fields to flush out rats to catch them and consume. They live mostly in eastern India extending themselves to central and northern parts till Nepal (Fig. 1). According to the elderly members of the Mushar community they have migrated from the East to different parts of the country and further to Nepal. It is interesting to note that in some of the Indian states they are considered as a scheduled caste (SC) and in some others as a scheduled tribe (ST) (Sharma 1998; Sinha 2002). They were hunters and gatherers initially but now live in villages and have adopted agriculture. Most of them currently speak Indo-European language but few (older people) are bilingual, speaking Indo-European (Bhojpuri, Maithili, Magahi and Bengali) and Austro-Asiatic (Munda) languages. For the comparative study, besides Mushar, we have also analyzed the same markers in contemporary populations holding the same social status. The details are given in Table 1.

### Methods and Data Analysis

5-10 ml of blood sample was collected from each of the 168 Mushar, 135 Austro-Asiatic and 151 Indo-European subjects, after obtaining consent from the respective populations. These samples were drawn covering Uttar Pradesh, Madhya Pradesh, Jharkhand and Bihar states. The details of population-wise samples and their geographical and linguistic affiliations are given in Table 1 and Figure 1.

All the samples have been analysed for the mtDNA control region as well as appropriate coding region markers (Primer numbers 1, 2, 3, 6, 15, 18, 22, 24 of Reider et al. 1998) in order to assign them in appropriate haplogroups. Y

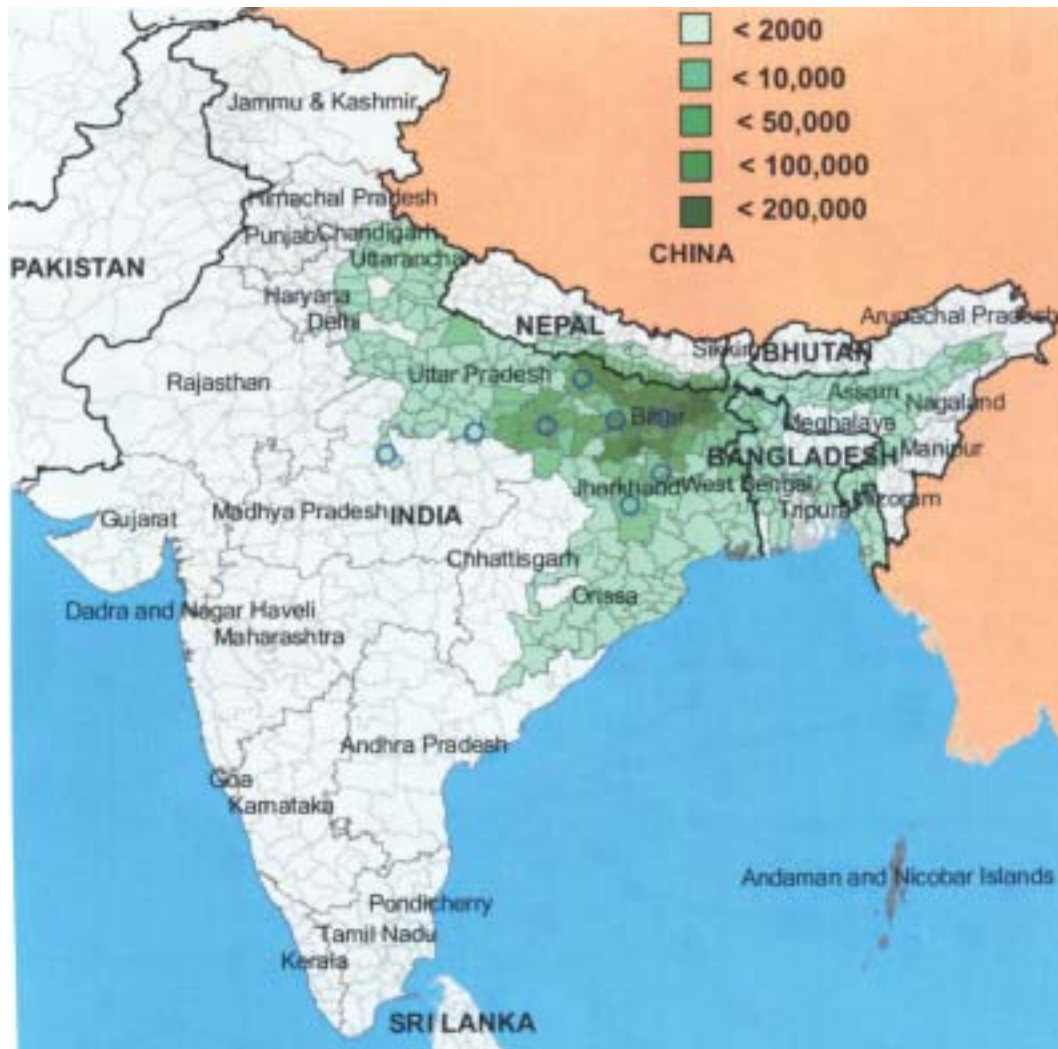


Fig. 1. The geographical distribution of Mushar population in South Asia. The sampling location is shown by circles and district wise Mushar population size is colour coded

chromosomal haplogroups were characterized with the help of 16 biallelic markers (M9, M45, M89, M130, M356, M52, M82, M304, M175, M95, M207, M124, M173, SRY1532, M20 and M76) typed by RFLP and sequencing methods. All the samples have been analysed in EBC, Tartu except the 5 Satnami and 14 Basor samples which have been analysed in CCMB, Hyderabad. Sequencing was carried out in ABI 3730 and 3730XL DNA analyzers. Individuals with ambiguous haplogroup affiliation were subjected to further sequencing of informative coding region stre-

tches. To minimize errors both strands were double-sequenced. Principal component analysis (PCA) of mtDNA as well as Y chromosome was performed as described earlier (Richards et al. 2002), using the software POPSTR provided by H. Harpending. Median joining network (Bandelt et al. 1999) was constructed using the Network 4.0 program (A. Rohl; Shareware Phylogenetic Network Software Web site) with default settings and finally drawn manually. Caste and tribal isofrequency maps were generated using the Kriging method in Surfer7 program of Golden

**Table 1: Linguistic and Geographical distribution of studied populations. AA-Austro-Asiatic, IE-Indo-European.**

<i>State</i>	<i>District</i>	<i>Status</i>	<i>Population</i>	<i>Lingustic affiliation</i>	<i>Sample</i>
Uttar Pradesh	Varanasi	Scheduled Tribe	Mushar	AA/IE	24
Uttar Pradesh	Gorakhpur	Scheduled Tribe	Mushar	AA/IE	21
Uttar Pradesh	Varanasi	Scheduled Caste	Harijan	IE	20
Uttar Pradesh	Varanasi	Scheduled Caste	Bhar	IE	22
Bihar	Bhojpur	Scheduled Tribe	Mushar	AA/IE	21
Bihar	Begusarai	Scheduled Tribe	Mushar	AA/IE	22
Bihar	Bhojpur	Scheduled Caste	Dusadh	IE	22
Bihar	Begusarai	Scheduled Caste	Lohra	IE	16
Bihar	Begusarai	Scheduled Tribe	Ho	AA	25
Bihar	Bhojpur	Scheduled Tribe	Munda	AA	20
Jharkhand	Ranchi	Scheduled Tribe	Mushar	AA/IE	18
Jharkhand	Bokaro	Scheduled Tribe	Mushar	AA/IE	21
Jharkhand	Bokaro	Scheduled Caste	Cheek-Baraik	IE	20
Jharkhand	Ranchi	Scheduled Tribe	Santhal	AA	23
Jharkhand	Ranchi	Scheduled Tribe	Munda	AA	22
Madhya Pradesh	Sidhi	Scheduled Tribe	Mushar	AA/IE	21
Madhya Pradesh	Bhopal	Scheduled Tribe	Mushar	AA/IE	20
Madhya Pradesh	Bhopal	Scheduled Tribe	Kharia	AA	21
Madhya Pradesh	Sidhi	Scheduled Tribe	Mawasi	AA	24
Madhya Pradesh	Sidhi	Scheduled Caste	Kanwar	IE	19
Madhya Pradesh	Sidhi	Scheduled Caste	Satnami	IE	18
Madhya Pradesh	Bhopal	Scheduled Caste	4Basor	IE	14

Software. Analysis of Molecular Variance (AMOVA) was also carried out for both Y-SNP and STR data using ARLEQUIN 3.01 software (Excoffier et al. 2005).

## RESULTS AND DISCUSSION

Several tribes have adopted other languages from juxtaposed communities (Diamond and Bellwood 2003). The hypothesis that Mushar have shifted their language from Austro-Asiatic to Indo-European has been well supported (Opler and Shukla 1968; Sharma 1998; Mishra 2001; Sinha. 2002). However, the genetic data have not been hitherto examined to determine if they had only shifted their language, not the genes. The mtDNA data analysis of Mushar population revealed a haplotype sharing with both Austro-Asiatic and Indo-European speakers. However, they share most of the haplotypes with Austro-Asiatic populations (Fig. 2). The mtDNA pool of the Mushar population comprises of several indigenous haplogroups which are typical to South Asia. The Austro-Asiatic (Mundari) specific haplogroups such as M40a, M45, R7 and R6a are also more frequent in the Mushar population and these are exclusively shared with the surrounding Austro-Asiatic populations considered for this study (Fig. 2). Most of the Indo-European samples don't share haplotypes with either the surround-

ing Austro-Asiatic or Mushar populations. The Y SNP tree also shows that as in the case of neighbouring Austro-Asiatic populations, O2a (M95) is the most frequent haplogroup in the Mushar populations (Fig. 3).

To determine whether the overall gene pool of Mushar is truly closer to Austro-Asiatic rather than to Indo-European-speaking populations in South Asia, we have constructed the haplogroup frequency based PCA plot for both mtDNA and Y chromosome markers (Fig. 4). Indeed, this analysis shows unambiguously that the Mushar population clusters with the Austro-Asiatic populations both in the mtDNA and Y chromosomal PCA plots. The major contributor haplogroups of mtDNA PCA plot of Mushar and Austro-Asiatic cluster are M40a, M45, R7 and R6a. The Analysis of Molecular Variance (AMOVA) based on mtDNA and Y-SNPs (Table 2) suggests that the Mushar is quite significantly differentiated from the Indo-Europeans ( $F_{ST}=0.082$  and  $0.353$  for mtDNA and Y chromosome, respectively), almost four-fold greater than its differentiation from the

**Table 2: The Analysis of Molecular variance (AMOVA) result**

	<i>mtDNA</i>	<i>Y-Chromosome</i>
	<i>Fst</i>	
Mushar VS Austro-Asiatic	0.025	0.082
Mushar VS Indo-European	0.082	0.353

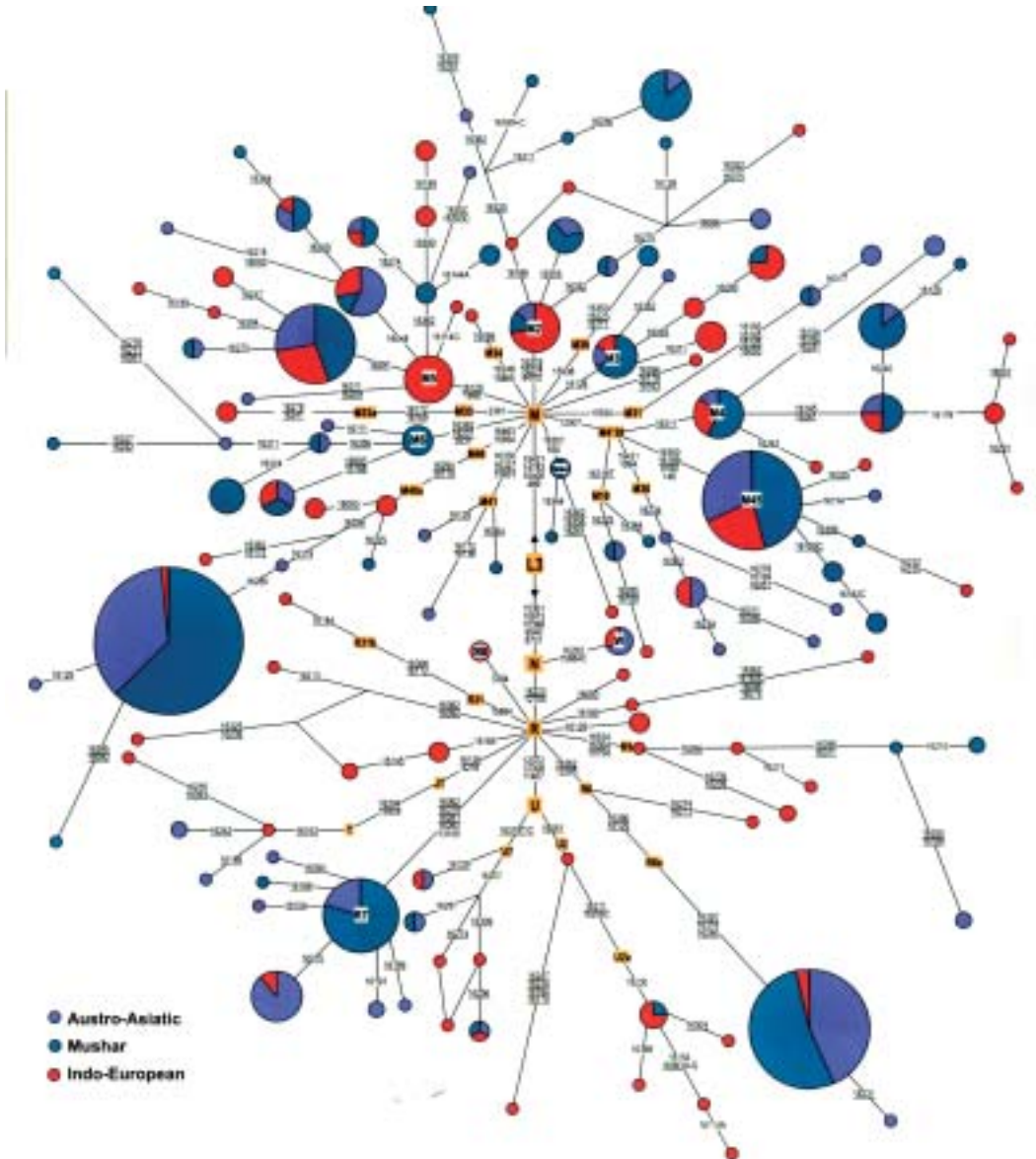


Fig. 2. The mtDNA network diagram of Mushar population with neighbouring Austro-Asiatic and Indo-European populations. This tree was drawn manually taking into account coding region as well as control region substitutions. Circle sizes are proportional to number of samples in the haplotype. Recurrent mutations are underlined.

Austro-Asiatic populations ( $F_{ST}=0.025$  and  $0.082$  for mtDNA and Y-chromosome, respectively).

We have also compared the haplogroup sharing for mtDNA and Y chromosome between all the three groups (Mushar, Austro-Asiatic and Indo-European). Both in the case of mtDNA and

Y chromosome, the Mushar and Austro-Asiatics share the common genetic features when compared with the Indo-European speakers (Fig. 5). It has been suggested that a caste having the same name may have different genetic makeup at different geographical regions (Karve and

Malhotra 1968; Majumder 2001). To test this result in the tribal population we compared the haplotype sharing among Mushar sub-populations collected from different geographical areas (Table 1, Fig. 1, 5). The largely uniform distribution of different haplogroups of mtDNA and Y chromosome suggest a common paternal and maternal origin of

these groups even though they inhabit distant geographic regions.

The most probable explanation of language shift in Mushars from Austro-Asiatic to Indo-European is the loss of their traditional lifestyle as hunters-gatherers in a surrounding, where the agricultural land has been the property of Indo-

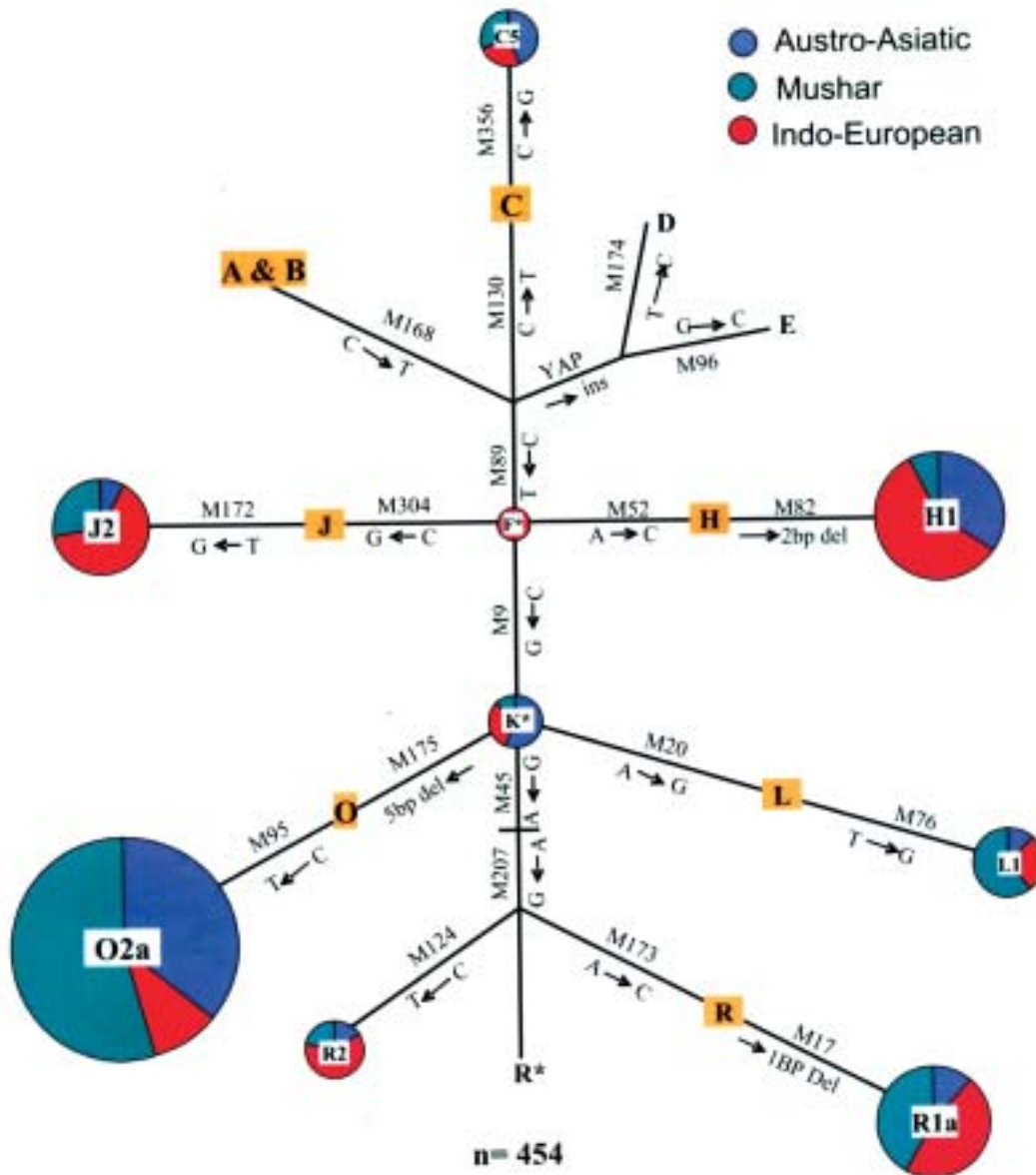


Fig. 3. The most parsimonious tree of Y chromosomal haplotypes of studied populations

European speaking communities. As they had to work in the fields owned by Indo-European-speaking people, it became advantageous to learn the language of the land-owners, possibly also linked to enhanced social prestige and desire for socioeconomic advancement. Equally importantly, children started to receive school education in an Indo-European language, facilitating gradual language shift over the generations.

To conclude, our study revealed close genetic

affinity among geographically divided sub-populations of Mushars who in turn share a common gene pool with neighbouring Austro-Asiatic populations as a predominant characteristic of their haploid genetic background. Hence, in this particular case study on Mushars, one observes that their genetic heritage is in accord with the data provided by independent historical and linguistic knowledge. However, we notice that there are numerous cases, where historical

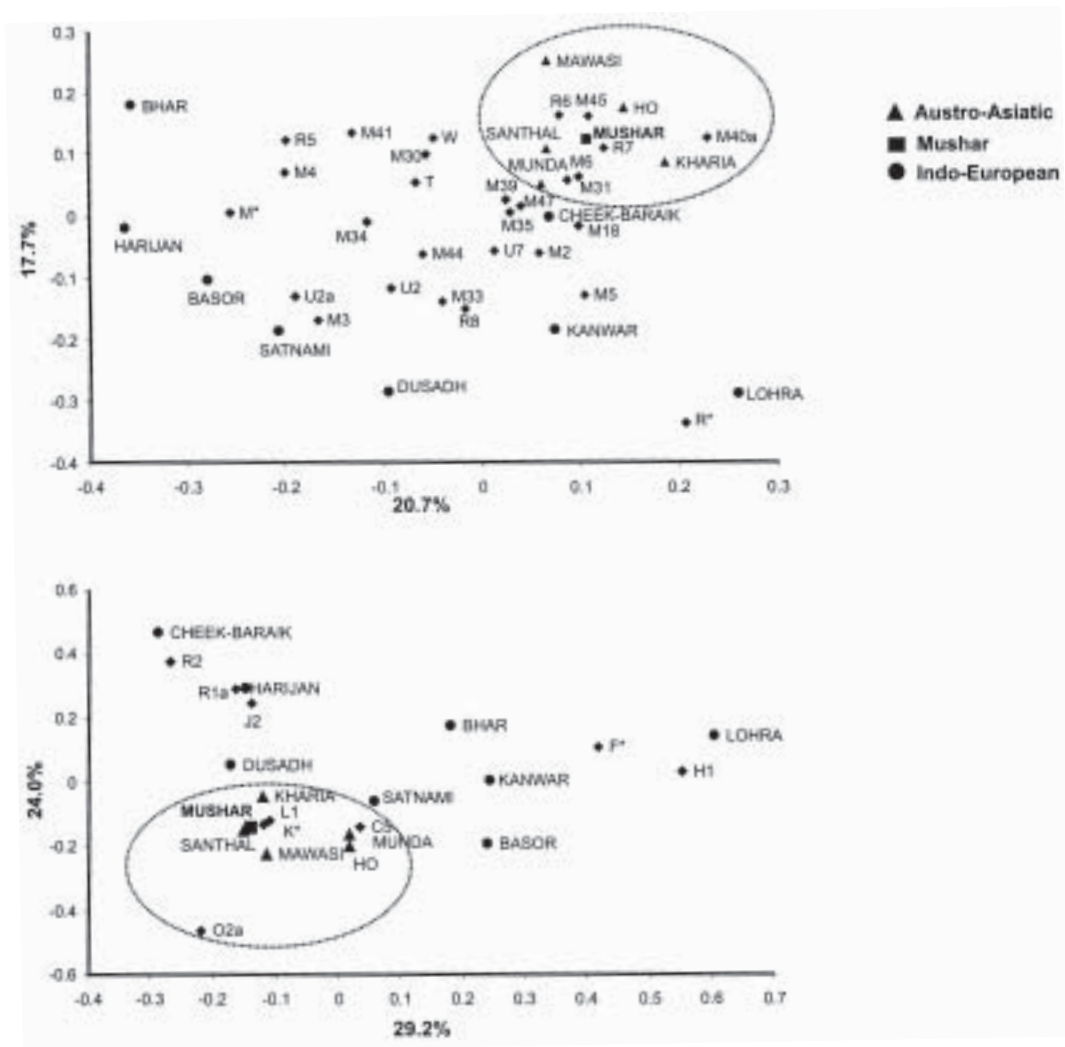


Fig. 4. The Principle component (PC) analysis based on mtDNA (upper pan) and Y chromosome (lower pan) haplogroups frequencies in South Asian studied populations. Notably, the Mushar population is clustered with Austro-Asiatic (Mundari) speaking populations.

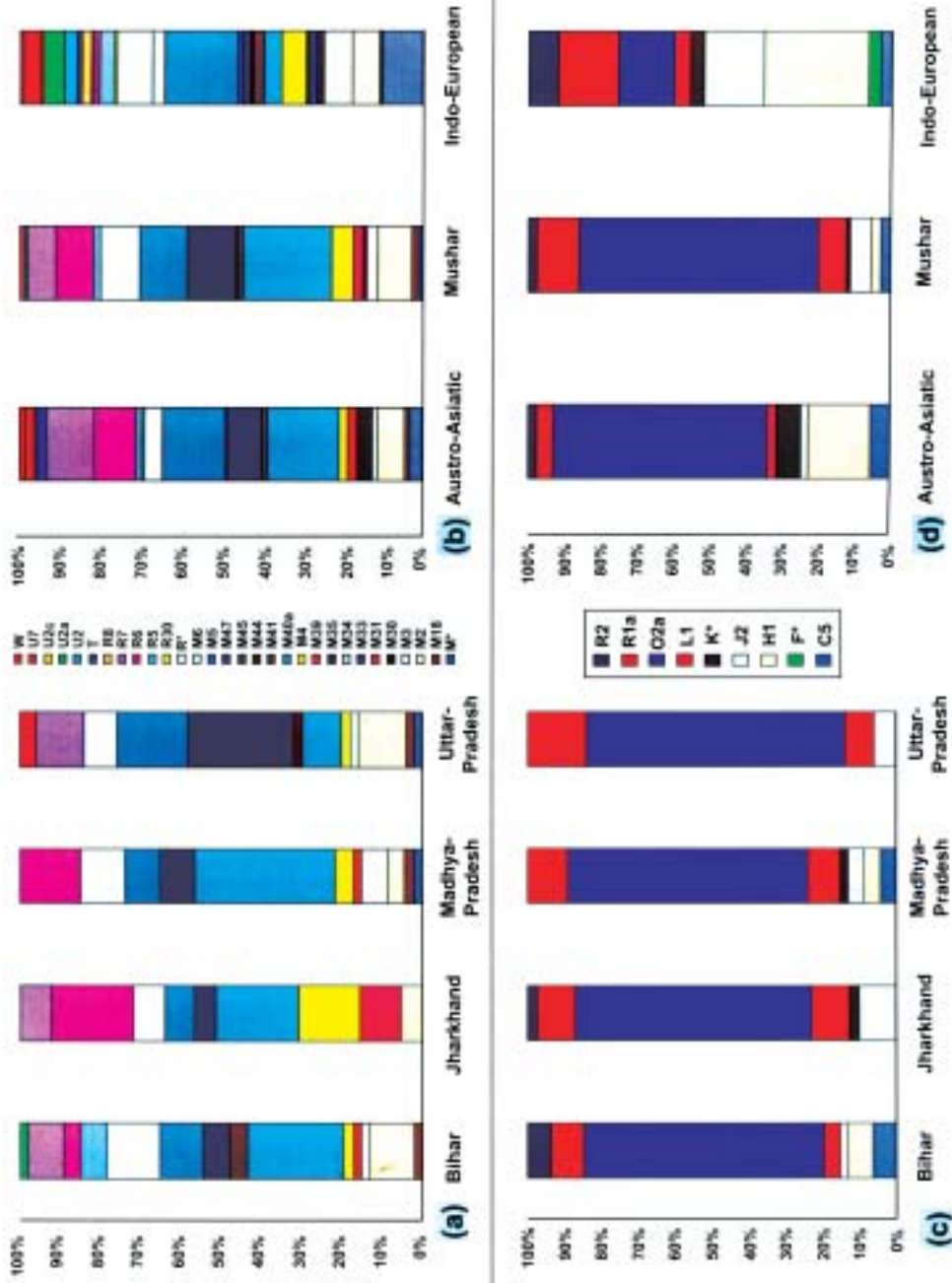


Fig. 5. Haplogroup sharing between different populations groups. The left upper and lower pan (a,c) show the mtDNA and Y-chromosome frequency distribution among different Mushhar sub-populations respectively. The right upper (b, d) and lower pan depict the mtDNA and Y-chromosome haplogroup frequency comparison of Mushhar with contemporary Indo-European and Austro-Asiatic groups.



linguistic knowledge is either absent or doubtful; for example, Oraon, Tharu, Lodha, Savara, Bonda, Mahali and several other populations of India which are of unclear or disputed linguistic origin. The analysis presented here suggests that recent rapid progress in the understanding of haploid genetics of the speakers of all major linguistic families in South Asia may offer insights helpful in the reconstruction of their putative past linguistic affiliation.

### ACKNOWLEDGEMENTS

We thank Jaan Lind, Ille Hilpus and Tuuli Mets for technical assistance. This work was supported by Estonian Basic Research grant SF0182474 (to RV) and Estonian Science Foundation grant 5574 (to TK). KT and LS supported by CSIR, India.

### REFERENCES

- Bandelt HJ, Forster P, Rohl A 1999. Median joining networks for inferring intraspecific phylogenies. *Mol Biol Evol*, **16**: 37-48.
- Cavalli-Sforza LL, Menozzi P, Piazza A 1994. *The History and Geography of Human Genes*. Princeton, NJ: Princeton Univ. Press.
- Chaubey G, Metspalu M, Kivisild T, Villems R 2007. Peopling of South Asia: investigating the caste-tribe continuum in India. *Bioessays*, **29**: 91-100.
- Cordaux R, Saha N, Bentley GR, Anuger R, Sirajuddin SM, Stoneking M 2003. Mitochondrial DNA analysis reveals diverse histories of tribal populations from India. *Eur J Hum Genet*, **11**: 253-64.
- Diamond J, Bellwood P 2003. Farmers and their languages: the first expansions. *Science*, **300**: 597-603.
- Excoffier L, Laval G, Schneider S 2005. An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**: 47-50.
- Ingman M, Kaessmann H, Paabo S, Gyllensten U 2000. Mitochondrial genome variation and the origin of modern humans. *Nature*, **408**: 708-13.
- Karve I, Malhotra KC 1968. A Biological Comparison of Eight Endogamous Groups of the Same Rank. *Current Anthropology*, **9**: 109-124.
- Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis K, Passarino G, Underhill PA, Scharfe C, Torroni A, Scozzari R, Modiano D, Coppa A, de Knijff P, Feldman M, Cavalli-Sforza LL, Oefner PJ 2006. The role of selection in the evolution of human mitochondrial genomes. *Genetics*, **172**: 373-387.
- Kivisild T, Rootsi S, Metspalu M, Mastana S, Kaldma K, Parik J, Metspalu E, Adojaan M, Tolk HV, Stepanov V, Golge M, Usanga E, Papiha SS, Cinnioglu C, King R, Cavalli-Sforza L, Underhill PA, Villems R 2003. The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *Am J Hum Genet*, **72**: 313-332.
- Kivisild T, Papiha SS, Rootsi S, Parik J, Kaldma K et al. 2000. An Indian Ancestry: A Key for Understanding Human Diversity in Europe and Beyond In: C Renfrew C, K Boyle (Eds): *Archaeo-genetics: DNA and the Population History of Europe*. Cambridge: McDonald Institute for Archaeological Research, pp. 267-75.
- Kivisild T, Kaldma K, Metspalu M, Parik J, Papiha S, Villems R 1999. The place of the Indian mtDNA variants in the global network of maternal lineages and the peopling of the Old World. In: R, Deka, SS Papiha (Eds.), *Genomic Diversity*. New York: Kluwer Academic/Plenum Publishers, pp. 135-52.
- Kulick D 1992. *Language Shift and Cultural Reproduction Socialization, Self and Syncretism in a Papua New Guinean Village*. Cambridge: Cambridge University press.
- Kumar V, Reddy AN, Babu JP, Rao TN, Langstieh BT, Thangaraj K, Reddy AG, Singh L, Reddy BM 2007. Y-chromosome evidence suggests a common paternal heritage of Austro-Asiatic populations. *BMC Evol Biol*, **7**: 47
- Majumder PP 2001. Indian Caste Origins: Genomic Insights and Future Outlook. *Genome Res*, **11**: 931-932.
- Metspalu M, Kivisild T, Metspalu E, Parik J, Hudjashov G, Kaldma K, Serk P, Karmin M, Behar DM, Gilbert MT, Endicott P, Mastana S, Papiha SS, Skorecki K, Torroni A, Villems R 2004. Most of the extant mtDNA boundaries in South and Southwest Asia were likely shaped during the initial settlement of Eurasia by anatomically modern humans. *BMC Genet*, **5**: 26.
- Mishra VN 2001. Prehistoric human colonization of India. *J of Bio*, **26** supp.
- Opler ME, Shukla S 1968. Palanquin Symbolism: The Special Vocabulary of the Palanquin-Bearing Castes of North Central India. *The Journal of American Folklore*, **81**: 216-234.
- Palanichamy M, Sun C, Agrawal S, Bandelt H-J, Kong Q-P, Khan F, Wang CY, Chaudhuri TK, Palla V, Zhang YP 2004. Phylogeny of mtDNA macrohaplogroup N in India based on complete sequencing: implications for the peopling of South Asia. *Am J Hum Genet*, **75**: 966-978.
- Reid, L. 1994. In: *Language Contact and Change in the Austronesian World*, T. Dutton, D. Tryon (Eds.) Berlin: Mouton de Gruyter, pp. 443-76.
- Richards M, Macaulay V, Torroni A, Bandelt HJ 2002. In search of geographical patterns in European mitochondrial DNA. *Am J Hum Genet*, **71**: 1168-74.
- Rieder MJ, Taylor SL, Tobe VO, Nickerson DA 1998. Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. *Nucleic Acids Res*, **26**: 967-973.
- Ross M 1997. In *Archaeology and Language* I, R. Blench and M. Spriggs, Eds. (Routledge, London, pp.209 – 261).
- Sahoo S, Singh A, Himabindu G, Banerjee J, Sitalaximi T, Gaikwad S, Trivedi R, Endicott P, Kivisild T, Metspalu M, Villems R, Kashyap VK 2006. A prehistory of Indian Y chromosomes: evaluating demic diffusion scenarios. *Proc Natl Acad Sci U S A*, **103**: 843-8.

- Sengupta S, Zhivotovsky LA, King R, Mehdi SQ, Edmonds CA, Chow CE, Lin AA, Mitra M, Sil SK, Ramesh A, Usha Rani MV, Thakur CM, Cavalli-Sforza LL, Majumder PP, Underhill PA 2006. Polarity and temporality of high-resolution Y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of central asian pastoralists. *Am J Hum Genet*, **78**: 202-21.
- Sharma JD. 1998. *Mushar: Bahu Bhasiya Samaj-in Hindi*. Thankur Prasad books.
- Sinha AN 2002. *The Mushar: A Socio-Economic Study*. Patna. Institute of Social Studies, India.
- Sun C, Kong Q-P, Palanichamy MG, Agrawal S, Bandelt H-J, Yao YG, Khan F, Zhu CL, Chaudhuri TK, Zhang YP 2006. The dazzling array of basal branches in the mtDNA macrohaplogroup M from India as inferred from complete genomes. *Mol Biol Evol*, **3**: 683-690.
- Thangaraj K, Chaubey G, Singh VK, Vanniarajan A, Ismail T, Reddy AG, Singh L 2006. In situ origin of deep rooting lineages of mitochondrial Macrohaplogroup 'M' in India. *BMC Genomics*, **7**: 151.
- Thangaraj K, Sridhar V, Kivisild T, Reddy AG, Chaubey G, Singh VK, Kaur S, Agarawal P, Rai A, Gupta J, Mallick CB, Kumar N, Velavan TP, Suganthan R, Udaykumar D, Kumar R, Mishra R, Khan A, Annapurna C, Singh L 2005. Different population histories of the Mundari- and Mon-Khmer-speaking Austro-Asiatic tribes inferred from the mtDNA 9-bp deletion/insertion polymorphism in Indian populations. *Hum Genet*, **116**: 507-17.
- Torrioni A, Achilli A, Macaulay V, Richards M, Bandelt HJ 2006. Harvesting the fruit of the human mtDNA tree. *Trends Genet*, **22**: 339-45.
- Wiik, K 2002. Eurooppalaisten juuret. Atena, Jyvaskyla.